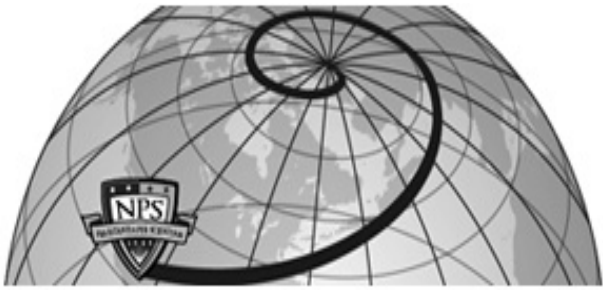

1953

Symmetry of physical laws. Part I : symmetry in space-time and balance theorems Symmetry in space-time and balance theorems.

Michael Satosi Watanabe

SYMMETRY OF PHYSICAL LAWS: PART I, SYMMETRY IN SPACE-TIME AND BALANCE THEOREMS.

UNITED STATES
NAVAL POSTGRADUATE SCHOOL

SYMMETRY OF PHYSICAL LAWS

Part I

Symmetry in Space-Time and Balance Theorems

-By-
Michael Satosi Watanabe,
Professor of Physics

Research Paper No. 2

SYMMETRY OF PHYSICAL LAWS

PART I

-Symmetry in Space-Time and Balance Theorems-

by

Michael Satosi Watanabe

Symmetry of Physical Laws

Part I

-Symmetry in Space-Time and Balance Theorems-

by Michael Satosi Watanabe

U.S. Naval Postgraduate School, Monterey, California

Abstract

In default of the theorem of "detailed balance": $P_{ij} = P_{ji}$, with regard to elementary transition probabilities, several "balance" theorems are introduced and proved on the basis of symmetry of physical laws in space-time. (1) First theorem of "averaged balance" (#5): We can establish $\overline{P}_{ij} = \overline{P}_{ji}$ by averaging over quantities of "minus class." Table V (#3) gives a list of "minus" quantities. (2) The so-called "detailed balance of collisions" in classical physics is a special case of Theorem (1). (3) Heitler-Coester's theorem of "semi-detailed balance" is also a special case of Theorem (1). (4) Second theorem of "averaged balance" (#5): We can establish $\overline{P}_{ij} = \overline{P}_{ji}$ by averaging over quantities with $\wp_R = -1$. The quantities with $\wp_R = -1$ are listed on Table II (#2). (5) Theorem of "cyclic balance" (#7): In classical physics, a chain of transitions $i \to j \to k \to \ldots \to i$ repeats itself cyclically. (6) Theorem of "long-range balance" (#7): The time average of transition probability from i to j is equal to the time average of transition probability from j to i. Theorems (1), (2) and (3) are direct consequences of inversibility (covariance for space-and-time inversion). Theorem (4) is a consequence of

reversibility (covariance for time reversal). Theorems (5) and (6) are connected with ergodicity of Markoff's chains. This ergodicity is proved by the condition of bilateral normalization of transition probabilities: $\sum_j P_{ij} = 1$, $\sum_i P_{ij} = 1$. This bilateral normalization in turn can be derived from either reversibility or inversibility. The limits of validity of all these balance theorems in actual applications are carefully examined in the text.

#1. Introduction

Since some time it has come to general attention that the principle of detailed balance by no means represents a universal rule in quantum physics.[1] Thus, thanks to Pauli's timely remark, the demonstration of the H-theorem which does not utilize the assumption of detailed balance has acquired a new importance.[2]

The transition probability P_{ij} from state S_i to state S_j has to obey, due to the very nature of probability,

$$\sum_j P_{ij} = 1 , \qquad P_{ij} \geq 0. \tag{1.1}$$

However, the inverse normalization:

$$\sum_i P_{ij} = 1 \tag{1.2}$$

is not self-evident.

That conditions (1.1) and 1.2) are sufficient for derivation of the H-theorem was demonstrated by Husimi[3] and Stückelberg,[2] and a simplified version of this proof was given by Pauli.[2]

---

1. J. Hamilton and H.W. Peng, Proc. Roy. Ir. Acad., A49, 197 (1944); W. Heitler, Quantum Theory of Radiation (Oxford University Press, London, 1944) second edition, p. 252.
2. E.C.G. Stückelberg, Helv. Phys. Acta, 25, 577 (1952)
3. Kodi Husimi, Theory of Probability and Statistics (Kawade Shobo, Tokyo, 1942, in Japanese) p. 277.

Obviously, the detailed balance condition:

$$P_{ij} = P_{ji} \tag{1.3}$$

allows one to deduce (1.2) from (1.1), but this is of course too restrictive a condition.

It is known that the bilateral normalization, (1.1) and (1.2), can be derived from the unitarity of transition matrix (S-matrix) in quantum physics. But in this paper, the bilateral normalization is considered in connection with the symmetry of physical laws in space-time. Namely, it will be pointed out that either reversibility (covariance for time-reversal) or inversibility (covariance for space-and-time inversion) is sufficient to deduce the bilateral normalization, without making use of the unitarity of transition matrices or of detailed balance. Reversibility or inversibility has indeed a very clear physical meaning and may be considered as a more basic physical principle than the unitarity of transition matrices, which is specifically a quantum-mechanical situation. In any event, in the quantum theory of elementary processes there never appear transition matrices which do not obey reversibility or inversibility. Therefore, reversibility or inversibility can be considered to be a sufficiently general rationalization of the bilateral normalization.

The principle of detailed balance (1.3) is sometimes resorted to in problems other than the H-theorem. Therefore it is worthwhile investigating its limits of validity. From Boltzmann's classical work, it is clear that this principle is intimately related to inversibility. Our investigation will show that if the physical system has inversibility, the

theorem of detailed balance can be re-established in a broader sense with the help of hypotheses of elementary disorder (or simply chaos hypotheses) with regard to the physical quantities of what will be called the "minus class." To the minus class belong regular tensors and first kind pseudo-tensors of odd ranks and second and third kind pseudo-tensors of even ranks. This classification of tensors will be discussed in detail in #2.[4]

This result clearly explains that the principle of detailed balance in classical physics is bound to utilize the chaos hypothesis with regard to the positions of molecules, which are minus variables. Heitler-Coester's so-called theorem of semi-detailed balance[5] is also a variant of the above-mentioned general rule. Rather inexactly expressed, this theorem of semi-detailed balance means that we can re-establish the detailed balance by averaging over the spin directions of particles. Since spin is a minus variable, it is natural that a chaos hypothesis is necessitated with regard to this variable. Although this theorem is particularly convenient for considerations in the perturbation theory, its domain of validity should not be over-estimated. In order to apply this theorem, we have to describe the particles only by plane waves and to describe the electromagnetic field, not by the field strengths, but by its sources. Otherwise, we need further averaging or chaos hypotheses regarding other variables of minus class. On the other hand, it is also not a general rule

---

4. S. Watanabe, Sci.Pap.Inst.Phys.Chem.Res.(Tokyo) 39, 157 (1941); S. Watanabe, Phys.Rev. 84, 1008 (1951).
5. W.Heitler, loc.cit.;W.Heitler, lecture notes, Ecole d'Eté de Physique Théorique, 1952; F.Coester, Phys.Rev. 84, 1259 (1951).

that each time the particles have an "internal" freedom, averaging or chaos hypothesis is needed regarding this freedom.

Boltzmann already noticed that even if the detailed balance does not hold, or in a rough usage of words, if the system in state S_i does not return to S_i after a double transition: $S_i \rightarrow S_j \rightarrow S_i$, there will be a chain of transitions: $S_i \rightarrow S_j \rightarrow S_k \rightarrow \ldots \rightarrow S_i$ by which the system will come back to the original state. This "cyclic balance", or "closed cycle of corresponding collisions" as Tolman[6] calls it, can be considered as a generalization of detailed balance. Stückelberg[2] pointed out, without proof, that the mechanism of his H-theorem is connected with cyclic balance.[7] It is obvious that cyclic balance is a manifestation of the ergodic nature of physical phenomena.

In the last section of this paper, we shall give a simplified version of the ergodic theorem, using only the hypothesis of bilateral normalization of transition probabilities. This will provide a general (though schematized) basis for the theorem of cyclic balance, without referring to collision processes.

This simplified ergodic theorem cannot directly be applied to the actual physical problems, on account of various simplifying conditions, which will be explained at an appropriate place. For instance, the number of states is assumed to be finite, which is not permissible for the

---

6. R.C.Tolman, The Principles of Statistical Mechanics (Oxford University Press, London, 1938) p. 114.
7. The author is indebted to Prof W. Pauli who in a private communication emphasized this point.

applications in classical physics. We ignore also the important notion of "macroscopic cell" on the energy shell. But on the the other hand, our ergodic theorem has the advantages not only of being very simple and mathematically rigorous, but also of exhibiting all the essential assumptions necessary for the deduction of the ergodic nature of transition probabilities. In particular, it will be shown that, the simplifying conditions being admitted, the bilateral normalization is the necessary and sufficient condition for the "ergodicity." The term ergodicity will be defined in accordance with the general ergodic theorem in physics. If reversibility or inversibility is taken as the foundation of the bilateral normalization, we can attribute ergodicity to the reversibility or inversibility of the physical laws. As in other versions of the H-theorem and the ergodic theorem in quantum physics, here also, the non-commutability of the Hamiltonian with the operators defining the states plays an essential role.

It is intended in the second Part to examine reversibility, reflectibility (symmetry in space) and inversibility of quantum field theory,[8] and to discuss their bearings on the interaction types and other allied problems. In this Part, these symmetry properties are formally defined and

---

8. As far as reversibility is concerned, the subject is fairly fully covered in the second paper quoted under footnote 4. See also the earlier works on this subject: E.P.Wigner, Göttinger Nachr. 546 (1932); S.Watanabe, Le Deuxième Théorème de la Thermodynamique et la Mécanique Ondulatoire (Hermann et Cie, Paris, 1935); S.Watanabe, Sci. Pap.Inst.Phys.Chem.Research (Tokyo) 31, 109 (1937).

assumed to exist when necessary. The classification of physical quantities into four "kinds" is explained in this paper in a fashion which may seem unduly elaborate. But this will prove to be instrumental not only for the discussion of the principle of semi-detailed balance but also for the discussion of reversibility, reflectibility and inversibility in general.

#2. "Kinds" and "Classes" of Tensors

In this section, we shall give the mathematical definitions of the four "kinds" and two "classes" of tensors. In the next section, we shall first introduce a formal method to determine the kinds of tensors representing various physical quantities, and then clarify the physical implication of this determination.

We consider the entire group of congruent transformations of coordinates:

$$x'^{\mu} = a^{\mu}_{\cdot\nu}\, x^{\nu}, \qquad (\mu, \nu = 1, 2, 3, 0) \tag{2.1}$$

which leave invariant

$$x_{\mu} x^{\mu} = g_{\mu\nu}\, x^{\mu} x^{\nu} = (x^1)^2 + (x^2)^2 + (x^3)^2 - (x^0)^2. \tag{2.2}$$

It is essential in the investigations involving inversions to use only the real coordinates, lest the connectivity of the Minkowski space may be altered.

The tensors $t^{\mu\nu\cdots}$ of the "regular" kind are defined by the transformation rule:

$$t'^{\mu\nu\cdots} = a^{\mu}_{\cdot\kappa}\, a^{\nu}_{\cdot\lambda} \cdots t^{\kappa\lambda\cdots}, \qquad \text{(regular)} \tag{2.3}$$

which we write for simplicity as

$$t' = \mathrm{T}\, t. \qquad \text{(regular)} \tag{2.4}$$

The pseudo-tensors $t^{\mu\nu\cdots}$ of the first, second and third kinds are defined respectively by[4]

$$t' = \sigma\, \mathrm{T}\, t, \qquad \text{(1st kind)} \tag{2.5}$$

$$t' = \sigma_t\, \mathrm{T}\, t, \qquad \text{(2nd kind)} \tag{2.6}$$

and

$$t' = \sigma_s\, \mathrm{T}\, t, \qquad \text{(3rd kind)} \tag{2.7}$$

where $$\sigma = \sigma_t \sigma_s = \frac{\partial(x'^1, x'^2, x'^3, x'^0)}{\partial(x^1, x^2, x^3, x^0)} , \tag{2.8}$$

$$\sigma_t = \sigma_s \sigma = \frac{\partial x'^0}{\partial x^0} \Big/ \left| \frac{\partial x'^0}{\partial x^0} \right| , \tag{2.9}$$

$$\sigma_s = \sigma \sigma_t = \frac{\partial(x'^1, x'^2, x'^3)}{\partial(x^1, x^2, x^3)} \Big/ \left| \frac{\partial(x'^1, x'^2, x'^3)}{\partial(x^1, x^2, x^3)} \right| . \tag{2.10}$$

The usual definition of ordinary tensors includes regular and second kinds, and that of pseudo-tensors includes first and third kinds, since only the "orthochronous" transformations are considered in their definition.

It is obvious from the above definition that the kind to which a product (with or without contraction) of two tensors belongs is determined by the following rules: (a) The product of two tensors of the same kind is a regular tensor. (b) The product of a regular tensor and a pseudo-tensor of a given kind belongs to the last-named kind. (c) The product of two pseudo-tensors of different kinds is a pseudo-tensor whose kind is different from either one of the two factors. (Table I). These rules are symmetrical regarding three tensors involved in the operation of multiplication.

	reg.	1	2	3
reg.	reg.	1	2	3
1	1	reg.	3	2
2	2	3	reg.	1
3	3	2	1	reg.

Table I. The kind to which a product of two tensors belongs.

The antisymmetric* tensor $r^{\mu\nu\cdots}$ of the rank n which is "complementary" to an antisymmetric* tensor $t^{\mu\nu\cdots}$ of the rank (4-n) is defined by

$$r = (1/24)\,\epsilon_{\mu\nu\kappa\lambda}\,t^{\mu\nu\kappa\lambda}\,, \tag{2.11}$$

$$r_{\mu} = (1/6)\,\epsilon_{\mu\nu\kappa\lambda}\,t^{\nu\kappa\lambda}\,, \tag{2.12}$$

$$r_{\mu\nu} = (1/2)\,\epsilon_{\mu\nu\kappa\lambda}\,t^{\kappa\lambda}\,, \tag{2.13}$$

$$r_{\mu\nu\kappa} = \epsilon_{\mu\nu\kappa\lambda}\,t^{\lambda}\,, \tag{2.14}$$

$$r_{\mu\nu\kappa\lambda} = \epsilon_{\mu\nu\kappa\lambda}\,t\,, \tag{2.15}$$

where the tensor ϵ is completely antisymmetric, and its component $\epsilon_{\mu\nu\kappa\lambda}\,(=-\epsilon^{\mu\nu\kappa\lambda})$ is +1 or -1 in <u>any</u> coordinate system, according as (μ,ν,κ,λ) is an even or odd permutation of (1,2,3,0). To satisfy this definition, ϵ must be a pseudo-tensor of the fourth rank of the first kind.

This complementary relation is reciprocal with regard to r and t.[9] The kinds of r, t and ϵ (1st kind) are related by the product rule. Thus, the second line of Table I will give the kind of r as dependent on the kind of t,. The scalar defined by (2.11) as complementary to ϵ itself is a regular scalar and has the value -1 in any coordinate system. This $\epsilon_{\mu\nu\kappa\lambda}$ should not be confused with a regular tensor $\eta_{\mu\nu\kappa\lambda}$ which is +1 or -1 according as (μ,ν,κ,λ) is an even or odd permutation of (1,2,3,0) in a particular coordinate system. Such a tensor changes the signs of its components by a transformation with $\sigma = -1$. The complementary scalar to η is then a pseudo-scalar of the first kind.

---

9. However, a little caution must be taken regarding the sign. For instance, $r^{12} = t^{30}$ but $t^{12} = -r^{30}$, also $r = t^{1230}$ but $t = -r^{1230}$ according to the above definition.

*. The modifier "antisymmetric" applies of course only to ranks higher than one.

If a physical quantity is to be expressed as a tensor component, we have to determine (a) the rank of the tensor, (b) the component which represents it and (c) the kind of the tensor. We assume in this work that (a) and (b) are already determined by the Lorentz transoformations in the narrow sense, except for the ambiguity due to the possible complementary representation.

To determine the kind of a tensor, it is sufficient to examine its behavior for time-reversal (hereinafter reversion):

$$x^1, x^2, x^3 \to x^1, x^2, x^3 \quad ; \qquad x^0 \to -x^0 , \tag{2.16}$$

and its behavior for total space-reflection (hereinafter mirage):

$$x^1, x^2, x^3 \to -x^1, -x^2, -x^3 \quad ; \quad x^0 \to x^0 . \tag{2.17}$$

For reversion or mirage, a component Q of any tensor will retain or change its sign, but its absolute value remains unchanged. We write for reversion

$$Q' = \rho_R Q \quad , \qquad \rho_R = +1 \quad \text{or} \quad -1 \ , \tag{2.18}$$

and for mirage

$$Q' = \rho_M Q \quad , \qquad \rho_M = +1 \quad \text{or} \quad -1 . \tag{2.19}$$

The four possible combinations of the values of ρ_R and ρ_M will lead to the classification into four kinds.

According to the definition of the four kinds of tensors, ρ_R and ρ_M are directly determined by the kind of the tensor and by the nature of the component in consideration. Table II and Table III list ρ_R and ρ_M for various components up to the second rank. In the designation of the nature of components in these tables, "space" means $\mu=1,2,3$ and "time" means $\mu = 0$.

rank	component	reg.	1st	2nd	3rd
scalar		+	−	−	+
vector	space	+	−	−	+
	time	−	+	+	−
tensor 2nd rank	space-space time-time	+	−	−	+
	space-time	−	+	+	−

Table II. The sign of ρ_R for various components as dependent on the kind.

rank	component	reg.	1st	2nd	3rd
scalar		+	−	+	−
vector	space	−	+	−	+
	time	+	−	+	−
tensor 2nd rank	space-space time-time	+	−	+	−
	space-time	−	+	−	+

Table III. The sign of ρ_M for various components as dependent on the kind.

For the combination of reversion and mirage (hereinafter total inversion or inversion):

$$x^1, x^2, x^3, x^0 \rightarrow -x^1, -x^2, -x^3, -x^0 , \tag{2.20}$$

we have $$Q' = \rho_I Q , \tag{2.21}$$

with $$\rho_I = \rho_R \rho_M . \tag{2.22}$$

For total inversion, the transformation matrix T of (2.4) through (2.7) becomes simply

$$T = +I \quad \text{or} \quad -I , \tag{2.23}$$

according as the rank of the tensor is even or odd. I in (2.23) means the identity matrix. The σ's are here

$$\sigma = +1 , \quad \sigma_t = -1 , \quad \sigma_s = -1 . \tag{2.24}$$

Definitions (2.4) through (2.7) show that the coefficient ρ_I is then simply the product of T (2.23) and one of the σ's (2.24). Thus we obtain a simple rule for ρ_I, which depends only on the rank and the kind of the tensor: ρ_I is positive for regular tensors and first kind pseudo-tensors of even ranks and for second and third kind pseudo-tensors of odd ranks. ρ_I is negative for regular tensors and first kind pseudo-tensors of odd ranks and for second and third kind pseudo-tensors of even ranks. All the quantities of the former group will form the "plus class", and all the quantities of the latter group the "minus class."

rank	reg.	1st	2nd	3rd
even	+	+	−	−
odd	−	−	+	+

Table IV. The class of a tensor determined by its rank and kind.

#3. Determination of the Kinds of Physical Quantities

It is a basic assumption of this entire work that any physical quantity can be represented as a component of a tensor of a certain kind. We are now going to introduce a set of formal preseriptions by which the kinds of various physical quantities can be determined according to the definition of each quantity. Admittedly, "definitions" of physical quantities and "physical laws" involving those quantities are hardly separable in many cases. As a result, one may raise an objection to the "proof" of reversibility etc. to the effect that the kinds of the physical quantities are determined in such a way that the reversibility, etc. may hold automatically. The point is however that the same physical quantities appear in various physical laws, and that it is meaningful to verify that there is no internal contradictions among these laws. In the following, we shall use as elementary a definition as possible of each physical quantity to determine its kind. The basic rules serving this purpose are as follows:

(a) The attributes of elementary particles, i.e., rest-mass, electric charge, mesic charge, mangitude of spin, are regular scalars.

(b) The proper-time differential ds is a pseudo-scalar of the second kind.

Rule (b) means that the sign of ds is determined by the sign of the time differential dt. Besides these rules, we notice that the transformations (2.16) and (2.17) do not change the sign of the operation $\iiint_{-\infty}^{\infty} dx^1 dx^2 dx^3$.

This means that, as far as the sings ρ_R and ρ_M are concerned, a physical quantity and its density behave in the same manner.

To begin with, we note that four-velocity dx^μ/ds is a 2nd kind pseudo-vector, since dx^μ is a regular vector while ds is a 2nd kind pseudo-scalar. By the regular invariance of intrinsic mass, the momentum of a particle mdx^μ/ds becomes then a second kind pseudo-vector. This warrants the positive-definite definition of energy, since $\rho_R = \rho_M = 1$ for the time component of a 2nd kind vector. Force $m\,d^2x^\mu/ds^2$ must be a regular vector, for dx^μ is a regular vector and ds stands here sqaured. The fact that the total charge of an elementary particle is a regular scalar necessitates that the current-density vector should be a 2nd kind pseudo-vector, since the 2nd kind is the only kind of vectors whose time components (charge-density in this case) have $\rho_R = \rho_M = 1$. The orbital angular momentum, being the product of a position-vector (regular) and a momentum-vector (2nd kind), should be represented by the space-space components of a 2nd kind pseudo-tensor. From Tables II and III, we see that the space-space components of a 2nd kind pseudo-tensor has the same values of ρ_R and ρ_M as the space component of a 1st kind pseudo-vector. This suggests that the spin-density of a particle, if expressed as a vector, should belong to the 1st kind. The magnetic moment density can be pictured as the product of charge (regular) and angular momentum (2nd kind); hence it must be represented as the space-space components of a 2nd kind pseudo-tensor. If it is represented as space-time components of a tensor,

this tensor must belong to the 3rd kind in virtue of the theorem of complementary tensors. By the definition of the electric field as the force on a charge, the ρ_R and ρ_M of the electric field must be the same as those of the space components of a regular vector which is force. Thus, we see from Tables II and III that electric field, if represented as the space-time components of a tensor, must belong to the 2nd kind.. If we know from the Lorentz transformation in the narrower sense that electric field and magnetic field build a tensor, the magnetic field should then be represented as space-space components of a 2nd kind tensor. An alternative representation of the electromagnetic field is, in virtue of the theorem of complementary tensors, such that the electric field and the magnetic field are respectively represented as space-space components and space-time components of a 3rd kind pseudo-tensor. If the magnetic pole strength is defined as the ratio of the force to the magnetic field, it must behave like a first kind scalar, since we have $\rho_R = 1$ and $\rho_M = -1$ for force and $\rho_R = -1$ and $\rho_M = 1$ for magnetic field.

The kind to which energy-momentum density tensor belongs can be determined by the requirment that its space-time components and time-time component should behave like the momentum-energy vector which is a 2nd kind vector. This classifies the energy-momentum density tensor as a regular tensor. From the relation between the energy-momentum density tensor and the Lagrangian density, it follows that the latter has to behave like the diagonal elements of the former. In other words it is a regular scalar.

The ranks and kinds of various physical quantities being thus determined, their classes follow immediately from the rule which is tabulated in Table IV. All these results together with some results which can easily be inferred are listed in Table V.

rank	kind	class	quantities
scalar	reg.	+	intrinsic mass, electric charge, mesic charge, magnitude of spin, Lagrangian
	1st	+	magnetic pole strength
vector	reg.	−	position-time, force
	1st	−	spin density
	2nd	+	four-velocity, momentum-energy, current-charge density, electromagnetic potentials, linear polarization of photon
tensor	reg.	+	energy-momentum density
	2nd	−	angular momentum (space-space), electric field (space-time), magnetic field (space-space), electric moment density (space-time), magnetic moment density (space-space), circular polarization of photon (space-space)
	3rd	−	All the quantities listed under 2nd class pseudo-tensor become 3rd class pseudo-tensors by the interchange of space-space components and space-time components

Table V. Classification of various quantities into kinds and classes.

The determination of the kinds to which belong various "internal" variables (i.e., other than position-time and energy-momentum) of a spinor field requires a further discussion in quantum field theory, which will be given in **Part II. Only some of** the results will be given here. g_R of these quantities have already been determined in a previous paper.[10] g_M of these quantities can be shown to be the same as in the c-number theory as determined by the transformation properties of spinors. Table VI lists the classification of these quantities, assuming that the two spinors appearing in each expression belong to the same transformation rule for reversion and mirage, i.e., to the same "kind" of spinors.[10]

	c-number theory		q-number theory	
	kind	class	kind	class
$\psi^\dagger\psi$	2nd	−	reg.	+
$i\psi^\dagger\gamma_\mu\psi$	2nd	+	2nd	+
$i\psi^\dagger\gamma_\mu\gamma_\nu\psi$	2nd	−	2nd	−
$i\psi^\dagger\gamma_5\psi$	3rd	−	1st	+
$i\psi^\dagger\gamma_5\gamma_\mu\psi$	3rd	+	1st	−
$\psi^\dagger\gamma_5\gamma_\mu\gamma_\nu\psi$	3rd	−	3rd	−

Table VI. The kinds and classes of various tensorial quantities built with two spinors of the same kind. The results for two spinors of different kinds can easily be inferred from this table using the definition of the four kinds of spinors[10]

---

10. The second paper quoted under footnote 4. **These topics will be discussed in detail in Part II.**

The kind of the pi-meson field is the same as the kind of its source, since the differentiation operator, if involved, is a regular vector. If the spinors representing the nucleons before and after the emission or absorption of a pi-meson are of the same kind, then the above table (under q-number theory) will give immediately the kind of the pi-meson field.* An inspection of Table VI will tell that a combination of scalar and vector types of interaction and a combination of pseudo-vector and pseudo-tensor types of interaction are not allowed. This "exclusion rule" of combination arises not from mirage but from reversion due to the change of ρ_R in q-number theory.[11] This change of ρ_R is exactly what is required to give to these quantities their respective physical meanings.[10] (See for instance that spin $i\psi^+\gamma_5\gamma_\mu\psi$ becomes a 1st kind vector).

We now proceed to introduce the notions of "reversed state", "miraged state" and "inverted state." The determination of kinds of the physical quantities given above is based essentially on a comparison of the two descriptions of the same physical phenomenon referring to two different coordinate systems related to each other by (2.16) or (2.17). The physical insight into the meahings of kinds can be obtained more easily by an alternative interpretation of the transformation (2.16) or (2.17), namely by considering two phenomena connected by this transformation described by the same coordinate system.

---

11. This exclusion rule of combinations holds also when the nucleons before and after the pi-emission (absorption) belong to different kinds of spinors. See also G. Lüders, ZS. f. Phys. 133, 325 (1952).

* This is true for neutral pi-mesons, the situation is slightly more complicated for charged pi-mesons. See Part II.

Two phenomena are said to be reversed phenomena of each other if, by suitably choosing the coordinate origin all the space coordinates involved in one phenomenon at any instant $t = x^0$ become the same as those involved in the other phenomenon at $-t$. It is hereby understood that the corresponding coordinates refer to the same physical entities, say the particles of the same attributes. The two states of the physical system, one referring to a phenomenon at t, the other referring to its reversed phenomenon at $-t$, are said to be reversed states of each other.

From this definition follows that in two mutually reversed states, the same particles have the same positions but the opposite velocities. This entails that the current in the reversed state should have the opposite sign, resulting in the opposite sign of magnetic field, etc. The rest of the argument then follows the same pattern as in the preceding determination of kinds of tensors. We can confirm in this manner that the invariance or change of the signs of the physical quantities in the reversed state is exactly the same as ρ_R determined in the foregoing. The reversed state S_R of a state S can now be re-defined as a state in which all the physical quantities with $\rho_R = 1$ have the same values as in S and all the quantities with $\rho_R = -1$ have the same absolute values but with the opposite signs.

Two phenomena are said to be miraged phenomena of each other if, by a suitable choice of the coordinate origin, all the space coordinates involved in one phenomenon at t are the negative of all the coordinates involved in the other phenomenon at the same instant t, whereby these

coordinates are supposed to refer to the same physical entities. The states of the physical system in such two phenomena at the same instant are said to be the miraged states of each other.

Comparing the consequences of this definition of the miraged states with the preceding determination of the kinds of tensors, we can re-define the miraged state S_M of a state S as a state in which all the physical quantities with $\rho_M = 1$ have the same values as in S and all the physical quantities with $\rho_M = -1$ have the same absolute values but with the opposite signs.

Two phenomena are said to be totally inverted phenomena of each other if, with a suitable coice of the space-time origin, all the space-coordinates involved in one phenomenon at t are the negative of the corresponding space-coordinates involved in the other phenomenon at $-t$. The two states compared here are totally inverted states of each other. The totally inverted state S_I of a state S can be defined as a state in which all the physical quantities with $\rho_I = 1$ have the same values as in S and all the physical quantities with $\rho_I = -1$ have the same absolute values but with the opposite signs.

#4. Reversibility, Reflectibility and Inversibility

Every closed system of physical laws must include a time-dependent law from which predictive statements can be deduced. Thus such a theoretical system should be capable of answering questions of the following type: What is the probability $P(S \rightarrow S'; t)$ of finding a physical system in state S' at the end of a period of time t if the system was found in state S at the beginning of this period? Such a probability will be simply called transition probability from S to S'.

If the description of the system by states S and S' is maximal, i.e., as detailed as allowed in principle, the prediction may be called a microscopic or dynamical prediction, while in other cases it is only statistical.* If the transition probability refers to a "statistical" prediction, we shall use the symbol W instead of P. We shall deal only with the dynamical probability P in this section. In classical physics, P is either 1 or 0, while in quantum physics we have only

$$0 \leq P \leq 1 .$$

In classical physics, a state is maximally defined if the values of all the independent physical quantities are furnished. In quantum physics, by a maximally defined state is meant a "pure state", or a quantum state, in contrast to a "mixture" (Gibbs ensemble or density matrix). Such a pure state may be considered as an eigenstate of a set of mutually commuting operators, representing a group of physical quantities, although in some cases these operators may be quite complicated.

---

* According to this usage of words, the ordinary transition probability in quantum physics from one quantum state to another should be qualified as microscopical or dynamical and not statistical. The "statistical" transition probability in quantum physics then refers to a transition of a system known to be in a Hilbert subspace to another Hilbert subspace, where the dimensions of subspaces are more than one.

Covariance for reversion, or reversibility, means that a process and its reversed process have the same probability, i.e., the transition probability from S to S' during t is equal to the transition probability from the reversed state S_R' of S' to the reversed state S_R of S during time t. Symbolically:

$$P(S \to S';\ t) = P(S_R' \to S_R\ ;\ t). \tag{4.1}$$

If a state S is characterized by the values Q of physical quantities, S_R is characterized by $\rho_R Q$. We write for brevity

$$S = \{Q\},\ S_R = \{\rho_R Q\},\quad S' = \{Q'\},\quad S_R' = \{\rho_R Q'\} \tag{4.2}$$

Covariance for mirage, or reflectibility, then means, in a similar symbolism,

$$P(S \to S';\ t) = P(S'_M \to S_M\ ;\ t) \tag{4.3}$$

$$\text{with } S = \{Q\},\quad S_M = \{\rho_M Q\},\quad S' = \{Q'\},\quad S_M' = \{\rho_M Q'\} \tag{4.4}$$

Finally, convariance for total inversion, or inversibility, means

$$P(S \to S';\ t) = P(S'_I \to S_I\ ;\ t) \tag{4.5}$$

$$\text{with } S = \{Q\},\quad S_I = \{\rho_I Q\},\quad S' = \{Q'\},\quad S'_I = \{\rho_I Q'\}\ . \tag{4.6}$$

There is a simple theorem which follows directly from this definition, on account of (2.22).

Theorem: If a physical system obeying a certain set of physical laws has any two of the three kinds of convariance, reversibility, reflectibility and inversibility, then it has also the third one.

For instance, suppose that a system enjoys reversibility and inversibility. First by reversibility, we have

$$P(\{Q\}\to\{Q'\};t) = P(\{\rho_R Q'\}\to\{\rho_R Q\};t), \tag{4.7}$$

and second by inversibility

$$P(\{\rho_R Q'\}\to\{\rho_R Q\};t) = P(\{\rho_I\rho_R Q\}\to\{\rho_I\rho_R Q'\};t). \tag{4.8}$$

Combining these two we obtain (4.3), for

$$\rho_M = \rho_I\rho_R \tag{4.9}$$

It should, however, be noted that it is quite possible that a physical system possesses only one of the three kinds of covariance. Physical systems which obey reflectibility but not reversibility (hence not inversibility) are familiar to us. Husimi, in a private communication, pointed out a rather peculiar mechanical example which has inversibility but not reversibility. Such a case is also quite conceivable.

It may also very well happen that a physical system does not obey (4.1) and (4.3) with the right signs of ρ_R and ρ_M but does obey them with wrong signs of ρ_R and ρ_M for some of the quantities, leading however to the right signs of $\rho_I = \rho_R\rho_M$. This statement is true for any permutation of ρ_R, ρ_M and ρ_I.

In classical physics, all the physical laws are written in terms of tensorial components. Therefore, if the physical quantities appearing in an equation belongs to the same kind, reversibility and reflectibility (hence also inversibility) are automatically guaranteed, since the existence of a solution representing a process will imply the existence of another solution representing the reversed or miraged process. As this situation is well-known, we shall limit ourselves to some remarks of general nature.

The mechanical laws are covariant for both reversion and mirage as

far as the force $m d^2x^\mu/ds^2$ is equated to a regular vector. The space components should then have $\rho_R = +1$, $\rho_M = -1$. The frictional force, say $-k(\mathbf{v}/|\mathbf{v}|)\mathbf{v}^2$ (Newtonian) has $\rho_M = -1$ but $\rho_R = -1$, thus it satisfies reflectibility but not reversibility. The Lorentz force $e(\mathbf{E} + [\mathbf{v} \times \mathbf{H}])$ however, has the right signs. The Maxwellian equations are covariant for both reversion and mirage, since the electromagnetic field tensor and its source, the current vector, belong both to the second kind. (See Table V.) In contrast to this, Ohm's law, $\mathbf{I} = \sigma \mathbf{E}$ is not covariant for reversion: the left side has $\rho_R = -1$, $\rho_M = -1$ while the right side has $\rho_R = +1$, $\rho_M = -1$. (See Tables II, III, V).

In quantum physics, the physical laws are not written in tensorial expressions. Therefore we have to examine whether we can construct the whole theory in such a way that the expectation values (including eigenvalues) of all the relevant physical quantities behave for reversion and miarage as their respective kinds will dictate. The kinds of the physical quantities which have classical analogues can easily be determined by the classical physics. The purely quantum mechanical quantities such as spin, magnetic moment, etc., can also be determined by their relations to known classical quantities. For instance, from the conservation law of total angular momentum, we have to assume the same ρ_R and ρ_M for spin as for orbital angular momentum. This is what has been done in #3. The question as to whether quantum physics in its entirety can be formulated in a covariant way for reversion and mirage will be studied in Part II.

In the remainder of this paper it is assumed that the physical laws governing the physical systems under consideration obey reversibility, reflectibility and inversibility. This assumption may be considered to be warranted as far as atomistic laws are concerned. In particular, it is understood that these covariance properties exist independently of how the states S and S' are defined, in so far as they are maximally defined.

#5. Detailed Balance, Semi-Detailed Balance and Averaged Balance

The theorem of detailed balance, literally taken, would mean

$$P(S \to S' \ ; \ t) = P(S' \to S \ ; \ t). \tag{5.1}$$

This type of theorem, however, holds only in the first order perturbation theorem in quantum theory and is of course not of a general validity.

The so-called theorem of detailed balance in classical physics by no means claims (5.1), which is a dynamical or microscopic law, but it represents a statistical law in which the state of a system is characterized by a distribution function in velocities (or momenta). It may be written, to exploit the distinction between P and W, as

$$W(S \to S' \ ; \ t) = W(S' \to S \ ; \ t) \tag{5.2}$$

where S is defined only by a certain distribution in velocities. This theorem is based essentially on an interesting fact that, as far as the linear momentum and energy of particles are concerned, a state and its totally inverted state are identical, since $\rho_I = 1$ for all the four components of momentum-energy vectors which are pseudo-vectors of the second kind. (See Table IV).

This indicates that the basic fact underlying the theorem of detailed balance is the theorem of total inversibility:

$$P(S \to S' \ ; \ t) = P(S'_I \to S_I \ ; \ t), \tag{5.3}$$

which means that the (dynamical) transition probability from state S to state S' is equal to the (dynamical) transition probability from the inverted state S'_I of S' to the inverted state S_I of S, whereby the

inverted state is constructed from the original state by keeping the values of all the quantities of the "plus" class and changing the signs of all the quantities of the "minus" class. See Table V. It should be noted that since the position is a minus quantity, comparison must be made between a quantity at x in S and the same quantity at $-x$ in S_I. For instance the electric field (a minus quantity) at x in S must be equal in magnitude and opposite to the electric field at $-x$ in S_I. This is in agreement with the situation created by the source point of this electric field placed at x in S and the same source point placed at $-x$ in S_I.

If we now disregard the electromagnetic field strengths and do not use the notion "force" and consider only particles without spin, then the only difference between a state and its inverted state lies in position coordinates and angular momenta. (Table V). This means that positions should be miraged, the change in sign of angular momentum ensuing automatically from this. This mirage of coordinates ($x \rightarrow -x$) involves not only the mirage of positions of particles but also the mirage of shapes of particles and mirage of the boundary.

It is now clear that, if the molecules are spherical (or points) and spinless and the boundary is symmetrical with regard to mirage, then the classical theorem of detailed balance (5.2) can be deduced, without discussion of collision processes, simply by the assumption that the distribution function is independent of position. This exposes that the classical theorem of detailed balance is based on the "chaos" hypothesis with

regard to the positions of molecules. (Chaos in angular momenta is a result of the chaos in positions).

If the molecules are not spherical and/or the boundary is not mirage-invariant, the chaos hypothesis regarding positions does not guarantee (5.2), since we still have miraged molecules and the miraged boundary in the inverted state. If the particles have at least one plane of symmetry, the miraged shapes of the particles can be reached by some rotations. If the boundary has at least one plane of symmetry, the miraged boundary can be considered as a rotated position of the same boundary. In this case, we can further introduce a chaos hypothesis with regard to the orientations of the molecules. This will secure a type of relation (5.2), but the right hand side will still refer to the rotated position of the boundary. Only if the boundary has a symmetry with regard to mirage, we can have the classical theorem of detailed balance on assumption of two kinds of chaos, one regarding positions, the other regarding the orinetations of molecules. This situation explains why the usual illustrations of breakdown of the theorem of detailed balance concerns either non-spherical molecules[12] or some irregular boundary.[13] The first categories of cases of breakdown can be remedied by the assumption of chaos

---

12. R.C.Tolman's textbook (quoted under footnote 6). p.119. Heitler's lecture note (quoted under footnote 5). J.M.Blatt and V.F.Weiskopf, Theoretical Nuclear Physics (John Wiley & Sons, New York, 1952)p. 530.

13. E.H.Kennard, Kinetic Theory of Gases (McGraw-Hill, New York, 1938)p.57

with regard to the orientations of molecules if they have at least one plane of symmetry. It is possible that this hypothesis should be sufficient even if the molecules do not have a plane of symmetry, but in such a case we shall have to discuss the collision process more in detail. In case of a boundary which does not have a mirage-invariant shape we had better resort to a long-range balance, or cyclic balance, which we shall touch upon in the last section.

The above derivation of the classical theorem of detailed balance suggests an immediate generalization. We can always establish a statistical balance of type (5.2), by assuming as many chaos hypotheses as we need quantities of minus class to describe the state*. This general rule will be called "principle of averaged balance."

In classical physics as well as in quantum physics, we can limit the number of necessary minus quantities to a certain degree by suitably choosing the employed variables. For instance, we can use the electro-magnetic potentials (plus quantities) instead of electro-magnetic field strengths (minus quantities), or even describing the electromagnetic field only by its source. In more elementary examples, "force" can be replaced by "potential."

In quantum physics, there is further a simplifying situation due to the existence of non-commuting quantities. If we characterize particles by their momenta (plus), we have to disregard their positions (minus) and angular momenta (minus). Hence, for the quantum mechanics of spinless partilces, we have even the detailed balance in the strictest sense (5.1)

---

* We should keep in mind that in some cases the chaos hypothesis regarding one variable automatically entails the chaos hypothesis regarding another variable. For instance, the chaos regarding position will result in chaos regarding angular momentum. The same is true for spin and magnetic moment.

if we adopt the plane-wave description, provided of course that the total Hamiltonian satisfies inversibility.

In case of the particleswith spin, we can derive the statistical balance (5.2) only by one chaos hypothesis regarding spin. This can be done by describing the electromagnetic field by its sources and describing these charged particles by linear momenta. From Table V, we see that the remaining physical quantities of minus class are only spin and electromagnetic moments. Therefore, averaging over spin directions will yield the averaged balance (5.2). This is what is called the principle of semi-detailed balance by Heitler,[5] which is a special case of our principle of averaged balance.

The theorem of semi-detailed balance is obviously very convenient when applied to the usual perturbation theory in which plane waves are taken as unperturbed states, i.e., eigen-states of the non-interacting Hamiltonian. Moreover, the spin $i\psi^{\dagger}\gamma_5\gamma_a\psi$ (a = 1, 2, 3) in the direction of propagation of the plane wave and the magnetic moment $i\psi^{\dagger}\gamma_a\gamma_b\psi$ (a,b = 1,2,3) in a direction perpendicular to this direction of propagation commute with the non-interacting Hamiltonian in the Dirac theory as well as with the momentum operator. Therefore the characterization of a state by the momentum and one of these "internal" variables is suitable for discussions in the perturbation theory. But this is only one of the possible modes of description. For instance, the total angular momentum, which is a minus variable, is also a constant of motion of the non-interacting Hamiltonian. It should also be noted that the inverted state of a diverg-

ing wave is a converging wave, therefore it is also outside the scope of this theorem.

Furthermore, there is no reason to limit the initial and final states to the eigenstates of the non-interacting Hamiltonian. If, for instance, the probability of existence of particles is more or less localized (wave-packet), position variables (minus) will intervene in the description of state, thus we shall need again a hypothesis of chaos regarding these "external" variables. On the other hand, we could avoid internal variables of the minus group by using, for instance, the spin-orbit interaction energy instead of spin (or magnetic moment) itself. Indeed, if we consider two charged particles in interaction, the magnetic moment is reversed in the inverted state, but the magnetic field strength due to the other particle also changes its sign in the inverted state at the point, so that the magnetic interaction will remain unchanged. Of course, this description involves more or less localized particles, as a result of which we shall have to pay for this simplification by the use of a chaos hypothesis regarding "external" variables. Thus we are faced with a kind of complementarity; to avoid one kind of chaos, we have to introduce another kind of chaos. It should also be recalled that there are also internal variables of the plus group. Table VI.

We thus see that we have to be very cautious in application of Heitler-Coester's principle of semi-detailed balance. A general rule, which does not fail, is that we should first determine the classes (plus or minus) of all the variables used in the description of the system, and

that, if we want to use a theorem of type (5.2), we should assume chaos hypotheses for all the minus class variables (principle of averaged balance). We encounter sometimes in the existing literature, statements to the effect that every time particles have an internal freedom, we have to perform an averaging with regard to this freedom. But, the coincidence of external and internal variables with plus and minus variables in the case of Heitler-Coester's theorem is only accidental.

We have used the word averaging or chaos in the above exposition without clearly defining its meaning, which we now should like to do. We take a mixture M_1 of state S and its inverted state S_I with equal weights. Then the average probability (average over S and S_I) of transition of a system in this mixture to S' or S'_I is given by

$$W_{12} = \frac{1}{2}\{P(S \to S') + P(S \to S'_I) + P(S_I \to S') + P(S_I \to S'_I)\} \qquad (5.4)$$

This probability is, due to inversibility (5.3), equal to the average transition probability of a system in a mixture M_2 of S' and S'_I with equal weights to S or S_I:

$$W_{21} = \frac{1}{2}\{P(S' \to S) + P(S' \to S_I) + P(S'_I \to S) + P(S'_I - S_I)\} = W_{12} \qquad (5.5)$$

This is the exact meaning of (5.2).

It should not be understood that this general consideration provides any kind of justification for the chaos hypothesis (equal weight of a state and its inverted state). In fact, if a system in mixture M_1 had the same average transition probability to S' and to S'_I, i.e., if

$$\frac{1}{2}\{P(S \to S') + P(S_I \to S')\} = \frac{1}{2}\{P(S \to S'_I) + P(S_I \to S'_I)\} \qquad (5.6)$$

were true, then a "chaos" would remain a "chaos" after transition. But (5.6) is by no means guaranteed by inversibility. One general way of justifying the chaos hypothesis is to take a mixture of all the possible states with equal weight, then we can expect that a system in this general mixture has an equal average transition probability to S' and to S'_I. In this case, the equal weight of a state and its inverted state holds as well before as after transition. This property implies that $\sum_i P_{ij} = \sum_i P_{ik}$ where S_j and S_k are inverted states of each other. And this condition is guaranteed by the inverse normalization, which is always true whenever there is inversibility or reversibility, as we shall see in the next section.

As far as the theorem of averaged balance is concerned, averaging is supposed to be made over each pair of states S and S_I, or $\{Q\}$ and $\{\rho_I Q\}$. But in case of space-coordinates, pairing of $\mathbf{x}$ and $-\mathbf{x}$ with regard to a particular coordinate system does not have an invariant meaning for translation, therefore averaging over all the values of coordinates is usually required.

Our derivation of the theorem of averaged balance using inversibility shows that we can also introduce a second theorem of averaged balance by considering reversed states instead of inverted states. Thus, we can obtain a type of relation (5.2) by averaging, or introducing chaos hypotheses, with regard to all the variables involved which have $\rho_R = -1$ (instead of $\rho_I = -1$).

#6. Bilateral Normalization of Transition Probabilities

We shall first consider the condition of bilateral normalization of microscopic transition probabilities P, and second consider the same condition with regard to statistical transition probabilities W.

For a given maximally defined state S, we think of a series of maximally defined states $S_i (i = 1,2,3, \ldots)$ such that S is one of them, say, S_i. Let us take the physical system in any condition, and represent the probability of finding it in state S_i by p_i. If

$$\sum_i p_i = 1 \tag{6.1}$$

we speak of a "complete" set of states. While in quantum physics there are more than one such set, in classical physics there is one and only complete set. In classical physics, the number of possible values of i are usually continuously infinite, and even multi-dimensional, and p_i will be a kind of δ-function in these variables. Thus (6.1) should be understood as a schematic simplification of the situation.

Take two such complete sets (which may be the same or different) of states S_i and S_j and consider the transition probability:

$$P(S_i \to S_j;t) \tag{6.2}$$

Then, the definition of P (#4) results, in virtue of (6.1), in the normalization regarding the final states:

$$\sum_j P(S_i \to S_j;) = 1 \tag{6.3}$$

However, the normalization with regard to the initial states:

$$\sum_i P(S_i \to S_j;t) = 1 \tag{6.4}$$

is not guaranteed by the definition. We shall see that if the physical

system obeys reversibility or inversibility, the inverse normalization (6.4) follows from the first normalization (6.3). In the discussion which immediately follows, we shall only speak of reversibility, but the word "inversibility" can always be substituted for the word "reversibility."

The basis of the demonstration is the fact that if S belongs to a complete set S_i, its reversed state S_R also belongs to set S_i. Due to reversibility, if $S \rightarrow S'$ is a solution of the dynamical law, $S_R' \rightarrow S_R$ is also a solution, implying that if S is a possible state, S_R is also a possible state. In classical physics, there is one and only complete set of states, therefore, this means that if S is a member of the set, S_R is also its member.

In quantum physics, a pure state $S = \{Q\}$ can be considered as an eigen-state of a family of Hermitian operators which, though complicated at times they may be, represent some physical quantities belonging to one or other of the four kinds. Therefore, each of them has a definite sign of ε_R. If the reversed state is a possible state (which is the case here), this state must also be an eigen-state of this family of physical quantities, for it is characterized by $\{\varepsilon_R Q\}$ of the same physical quantities. Hence S_R belongs to the same complete set as S. We can thus conclude that a complete set of states is composed of self-reversed states and pairs of mutually reversed states.

This being the case, the summation over all the S_i's and the summation over all the S_{iR}'s must mean the same operation. First by reversibility,

$$\sum_i P(S_i \rightarrow S_j;t) = \sum_i P(S_{jR} \rightarrow S_{iR};t) \qquad (6.5)$$

and, due to the above remark,

$$= \sum_i P(S_{jR} \rightarrow S_i;t), \qquad (6.6)$$

which is on account of the first normalization condition (6.3) equal to unity. Hence (6.4). Q.E.D.

It is true that the physical meaning underlying the unitarity of transformation matrices in quantum physics is connected with the "completeness" of representation. But we have derived here the bilateral normalization without utilizing specifically quantum-mechanical relations.

We now pass to the bilateral normalization with regard to the statistical transition probabilities W. It is usually the case that states S defined only statistically exhibit also "completeness." In other words, we can consider any non-maximally defined state S as a member of a series of non-maximally defined states S_i ($i = 1,2,3,\ldots$) such that the probability w_i of finding the system in a state S_i of the series obey a normalization condition:

$$\sum_i w_i = 1 \qquad (6.7)$$

For example, in classical physics, after averaging over all space-coordinates, the state of each molecule of a gas is characterized only by velocities v. Then the velocity-space can be divided into small volume-elements, which certainly have the property of completeness in the sense of (6.7). We can also apply this consideration to a pair of molecules, as is usually done in the discussion of collision processes. In case of Heitler-Coester's mode of description, after the averaging over the spin-

directions, the possible values of the momenta will constitute a complete set.

Now, if we take such a complete set of non-maximally defined states, it is quite natural to assume that it includes the reversed (inversed) states of all its members. Then, we can conclude the bilateral normalization by the same argument as given above. Or, more simply, if such a set of non-maximally defined states is so chosen that the theorem of averaged balance (5.2) is true, then the inverse normalization:

$$\sum_{i} W(S_i \rightarrow S_j;t) = 1 \qquad (6.8)$$

follows immediately from the first normalization:

$$\sum_{j} W(S_i \rightarrow S_j;t) = 1 \qquad (6.9)$$

which is a consequence of (6.7).

It should be recalled that even in classical physics, the w's and W's are not limited to zero and unity, which is the case for the p's and P's in this form of physics.

#7. Ergodic Property of Transition Probabilities

It is well-known that the Markoff chain exhibits a particular property which may be called "ergodic." But the usual exposition of this subject is too mathematical in nature and often overly simplified by the assumption of detailed balance: $P_{ij} = P_{ji}$.

It is intended to point out in this section that the bilateral normalization of transition probabilities is just necessary and sufficient to derive the "ergodicity" of the Markoff chain, and to clarify in what sense we can here speak of an ergodic theorem. In the following, we shall discuss the subject in terms of the microscopic, or dynamical probabilities P, but we shall soon find that the main body of argument applies also to the statistical probabilities W.

We take a complete set $\mathfrak{T}$ of maximally defined states S_i, and the indices i, j, etc., of the S's are supposed always to refer to this same set. We limit ourselves to the cases where there are a finite number of states in the set:

$$i = 1,2,3,\ldots, r \ . \tag{7.1}$$

In classical physics, there are usually a continuously infinite number of the S's. In this case, the entire argument that follows offers only a mathematical model which may approximate the real physical situation. In quantum physics, (7.1) does not imply a real limitation, since we need actually consider only a limited region of energy values (micro-canonical shell) and we can also assume the space domain to be limited. Then the number of quantum states will become finite.

Among these states $S_i (i = 1,2,\ldots, r)$, some will be disconnected from one another due to various conservation laws. For instance, two states belonging to different values of the total angular momentum will allow no transitions from one to the other. Thus, the entire set of S_i will be divided into sub-sets, in each of which the states are "connected." Such a sub-set will be called hereinafter "sub-shell." A more rigorous definition of sub-shells will soon be given.

The theory of Markoff chains pertains to the "repeated" transition probability $P_{ij}^{(n)}$ which is defined by

$$P_{ij}^{(n)} = \sum_k P_{ik}^{(n-1)} P_{kj} = \sum_k P_{ik} P_{kj}^{(n-1)} \tag{7.2}$$

where

$$P_{ij}^{(1)} = P_{ij} = P(S_i \to S_j; \tau). \tag{7.3}$$

In classical physics, we have $P_{ij} = 1$ or 0, therefore also $P_{ij}^{(n)} = 1$ or 0. In quantum physics, $0 \leq P_{ij}^{(n)} \leq 1$. In both cases, we have

$$\sum_j P_{ij}^{(n)} = 1, \tag{7.4}$$

which follows from (7.2) in virtue of the first normalization:

$$\sum_j P_{ij} = 1. \tag{7.5}$$

In the same manner, the inverse normalization:

$$\sum_i P_{ij} = 1 \tag{7.6}$$

will result in

$$\sum_i P_{ij}^{(n)} = 1. \tag{7.7}$$

The classical physics is characterized by the fact that

$$P_{ij}^{(n)} = P(S_i \to S_j; n\tau), \tag{7.8}$$

which means that the physical system is not disturbed by observation.

In quantum physics, this is not the case in general, unless the operators defining S_i commute with the exact Hamiltonian. In quantum physics, the repeated transition probability (7.2) acquires a physical meaning only on assumption that the system is observed every τ seconds with the operators defining S_i. In other words, starting with a pure state S_i, we observe the system after τ seconds, and the result is statistically represented by a mixture (ensemble or density matrix) composed of various S_j with the weight P_{ij}. By repeating this process at each interval of τ seconds, we obtain after $n\tau$ seconds a mixture of S_j's with the respective weights $P_{ij}^{(n)}$. This means that although we start with the microscopic transition probabilities P_{ij}, we have to interpret the repeated transition probabilities $P_{ij}^{(n)}$ in quantum physics in terms of "mixtures." The ergodic theorem discussed in this section thus refers to a chain of repeated observations and should not be confused with the more important ergodic theorem[14] which refer to two observations, one at the initial instant and the other at the final instant.

Closely related to P_{ij}, and physically and mathematically more significant than these are the quantities:

$$\Omega_{ij}^{(n)} = \sum_{m=1}^{n} P_{ij}^{(m)}/n \qquad (7.9)$$

14. J. von Newmann, ZS. f. Phys. 57, 30 (1929); W.Pauli and M.Fierz, ZS. f. Phys. 106, 572(1937). For an interpretation of Neumann's ergodic theorem in terms of initial and final observations, see an article by S.Watanabe in the monograph: Louis de Broglie(Albin Michel, Paris, 1952) p. 385.

Physically important are they, since $\Omega_{ij}^{(n)}$ represents the "time average" of transition from S_i to S_j during the time $n\tau$ seconds, while $P_{ij}^{(n)}$ represents the transition probability at the instant $n\tau$ seconds after the initial instant. Indeed, the main concept in an ergodic theorem in physics is a comparison of "average in time" with "average in microcanonical ensemble." Mathematically useful are they, since $\Omega^{(n)}$ has a better convergence than $P^{(n)}$ for $n \to \infty$. Although $P^{(n)}$ is zero or unity in classical physics, $\Omega^{(n)}$ is not necessarily so: $0 \leq \Omega_{ij}^{(n)} \leq 1$.

From the first normalization (7.5) follows:

$$\sum_j \Omega_{ij}^{(n)} = 1, \tag{7.10}$$

and from the inverse normalization (7.6):

$$\sum_i \Omega_{ij}^{(n)} = 1. \tag{7.11}$$

Now the "ergodic theorem" which we are going to prove can be enunciated as follows: The time average $\Omega_{ij}^{(n)}(n \to \infty)$ of probability of finding a system, which started from any S_i, in a state S_j of the same sub-shell is equal to the a-priori probability of a state in the sub-shell, i.e., equal to $1/s$ if s is the number of states in the sub-shell. This statement is certainly a faithful adaptation of the general ergodic theorem to our simplified case, since the microcanonical ensemble represents a mixture of all the states on an energy shell with equal weights. The main purpose of this section is to show that the inverse normalization is the necessary and sufficient condition for this simplified ergodic theorem, the first normalization being always assumed for the Markoff chain.

We shall first enumerate without proofs some of the elementary

theorems and definitions regarding the Markoff chain which can be found in any exposition of the subject.[15] We shall denote by $\mathfrak{T}$ the original complete set of states (7.1). Only the first normalization is assumed in the following theorems.

Theorem I. The sequence:

$$\Omega_{ij}^{(1)}, \Omega_{ij}^{(2)}, \cdots \qquad (S_i, S_j \in \mathfrak{T}) \tag{7.12}$$

converges to a limit:

$$\lim_{n\to\infty} \Omega_{ij}^{(n)} = \Omega_{ij}^{\infty} . \tag{7.13}$$

Of course, we have

$$0 \leq \Omega_{ij}^{\infty} \leq 1 , \quad \sum_j \Omega_{ij}^{\infty} = 1 . \tag{7.14}$$

In the set $\mathfrak{T}$ there can be some states S_j such that the average transition probabilities Ω_{ij}^{∞} to them vanish for any arbitrary initial state S_i.

Definition I. The "vanishing" part $\mathfrak{N}$ of $\mathfrak{T}$ is the set of all states S_j such that

$$\Omega_{ij}^{\infty} = 0 . \quad (S_i \in \mathfrak{T}, \ S_j \in \mathfrak{N}) \tag{7.15}$$

For the rest of the original set: $\mathfrak{T}' = \mathfrak{T} - \mathfrak{N}$ (which can easily be shown not to be empty), we have the following theorem:

Theorem II. If $\Omega_{ij}^{\infty} > 0$, then $\Omega_{ji}^{\infty} > 0$. $(S_i, S_j \in \mathfrak{T}')$ (7.16)

In other words, if $\Omega_{ij}^{\infty} = 0$, then $\Omega_{ji}^{\infty} = 0$. $(S_i, S_j \in \mathfrak{T}')$ (7.17)

---

15. Theorems I through IV are given in Husimi's textbook (Husimi, op. cit.,[3] p. 280), but their physical applications in physics, including Theorems V, VI, are not given there. Husimi's exposition is based on K. Yoshida and S. Kakutani, Topological Mathematics (Iwanami, Tokyo, 1939, in Japanese) Vol. 2, p. 20. See also, W. Feller, Introduction to probability theory and its application, (John Wiley, New York, 1950) p. 307 ff.

Using Theorem II, we can divide $\mathfrak{r}'$ into sub-sets ("sub-shell") such that Ω_{ij}^{∞} is zero if S_i and S_j belong to different sub-shells, and $\Omega_{ij}^{\infty} > 0$ for S_i and S_j belonging to the same sub-shell.

Definition II.
$$\mathfrak{r}' = \mathfrak{E}_1 + \mathfrak{E}_2 + \cdots + \mathfrak{E}_p \tag{7.18}$$
$$\Omega_{ij}^{\infty} = 0 \quad (S_i \in \mathfrak{E}, S_j \in \mathfrak{E}') \tag{7.19}$$
$$\Omega_{ij}^{\infty} > 0 \quad (S_i, S_j \in \mathfrak{E}) \tag{7.20}$$

A sub-shell $\mathfrak{E}$ is disconnected from the vanishing part $\mathfrak{P}$ and from another sub-shell $\mathfrak{E}'$ not only in terms of Ω_{ij}^{∞} (See (7.15) and (7.19)) but also in terms of P_{ij}:

Theorem III.
$$\left.\begin{aligned} P_{ij} &= 0, \quad (S_i \in \mathfrak{E}, S_j \in \mathfrak{P}) \\ P_{ij} &= 0, \quad (S_i \in \mathfrak{E}, S_j \in \mathfrak{E}') \\ P_{ij} &= 0, \quad (S_i \in \mathfrak{E}', S_j \in \mathfrak{E}) \end{aligned}\right\} \tag{7.21}$$

Obviously the inverse of this theorem is not true. It can happen that $P_{ij} = 0$ even for S_i and S_j belonging to the same sub-shell, i.e., in spite of $\Omega_{ij}^{\infty} > 0$.

We now pass to study the properties of those Ω_{ij}^{∞} whose initial and final states belong to the same sub-shell $\mathfrak{E}$ consisting of s states S_i:
$$i = 1, 2, 3, \ldots, s. \tag{7.22}$$

In virtue of (7.21), we can derive from (7.4)
$$\sum_{j=1}^{s} P_{ij} = 1, \quad (S_i, S_j \in \mathfrak{E}). \tag{7.23}$$

Similarly, on account of (7.15) and (7.19), we have
$$\sum_{j=1}^{s} \Omega_{ij}^{\infty} = 1, \quad (S_i, S_j \in \mathfrak{E}). \tag{7.24}$$

Relations (7.23) and (7.24) show that the first normalization (7.4)

(7.14) remains unchanged when the initial and final states are limited to a sub-shell.

We are now prepared to interduce an important theorem:

Theorem IV. Ω_{ij}^{∞} $(S_i, S_j \in \mathfrak{E})$ is independent of the initial state S_i:

$$\Omega_{ij}^{\infty} = \Omega_j \qquad (S_i, S_j \in \mathfrak{E}). \tag{7.25}$$

Of course, we have on account of (7.24)

$$\sum_{j=1}^{s} \Omega_j = 1. \tag{7.26}$$

The discussion up to this point assumes only the first normalization. We now investigate the implication of the inverse normalization. In particular, we shall consider the condition:

$$\sum_{i=1}^{s} P_{ij} = 1, \quad (S_i, S_j \in \mathfrak{E}), \tag{7.27}$$

which exhibits a symmetry to (7.23). On account of (7.21), the summation with regard to S_k in

$$P_{ij}^{(n)} = \sum_k P_{ik} P_{kj}^{(n-1)}, \quad (S_i, S_j \in \mathfrak{E}\ ;\ S_k \in \mathfrak{T}),$$

actually extends only over $S_k \in \mathfrak{E}$. Hence

$$\sum_i P_{ij}^{(n)} = 1, \quad (S_i, S_j \in \mathfrak{E}), \tag{7.28}$$

and, by (7.9) and (7.13), also

$$\sum_i \Omega_{ij}^{(n)} = \sum_i \Omega_{ij}^{\infty} = 1, \quad (S_i, S_j \in \mathfrak{E}). \tag{7.29}$$

We now propose to show that the inverse normalization (7.27) is equivalent to the condition that Ω_{ij}^{∞} is not only independent of the initial state S_i (Theorem IV) but also independent of the final state S_j. This last condition can be written, in view of (7.26), as

$$\Omega_{ij}^{\infty} = 1/s \qquad (S_i, S_j \in \mathfrak{E}). \tag{7.30}$$

Theorem V. The necessary and sufficient condition for (7.30) is (7.27).

Proof: From the definition of Ω^{∞}_{ij} (7.9) (7.13), we can easily obtain

$$\Omega^{\infty}_{ij} = \sum_k \Omega^{\infty}_{ik} P_{kj} \quad (S_i, S_j \in \mathcal{E};\ S_k \in \mathcal{T}) \tag{7.31}$$

On account of (7.15) and (7.19), the summation over S_k, in reality, extends only over $S_k \in \mathcal{E}$. If (7.30) is the case, (7.31) becomes

$$\frac{1}{s} = \frac{1}{s} \sum_{k=1}^{s} P_{kj} \qquad (S_k, S_j \in \mathcal{E}) \tag{7.32}$$

showing that (7.27) is a necessary condition for (7.30).

Next we shall show that (7.27) is also a sufficient condition.

If (7.27) is true, then we have (7.29), which in view of (7.25) means

$$\sum_{i=1}^{s} \Omega^{\infty}_{ij} = s\,\Omega_j = 1 \qquad (S_i, S_j \in \mathcal{E}) \tag{7.33}$$

or

$$\Omega_j = \frac{1}{s} \qquad \text{Q.E.D.} \tag{7.34}$$

Theorem V is obviously equivalent to the ergodic theorem we stated at the beginning of this section.

It should be noted that our inverse normalization (7.27) is not necessarily equivalent to the inverse normalization with regard to the entire set:

$$\sum_i P_{ij} = 1, \quad (S_i, S_j \in \mathcal{T}). \tag{7.35}$$

It is however easy to see that if (7.35) is true then (7.27) is also true, and that if (7.27) is true and if the entire set $\mathcal{T}$ has no vanishing part $\mathcal{Q}$, then (7.35) is true. Actually, in (7.27), S_i and S_j can be extended, without any additional assumption, to all the states belonging

to $\gamma' = \gamma - \chi$:

$$\sum_{i} P_{ij} = 1, \quad (S_i, S_j \in \gamma'), \tag{7.36}$$

on account of (7.21). The summation in (7.36) extends to all the states in γ' . (7.36) is equivalent to (7.27).

Now, if (7.35) is true, we shall have (7.11), in which $S_i, S_j \in \gamma$. But this contradicts the existence of a vanishing part (7.15). Hence, if (7.35) is true, then $\gamma = \gamma'$, and (7.36) and (7.27) ensue. On the other hand, if we have (7.36) as a given premise, then the conclusion (7.35) can be drawn only with the help of an additional condition $\gamma = \gamma'$.

From whatever state one may start, there will be in the long run a vanishing probability of having the system in a state belonging to the vanishing part, (7.15), and also starting from a state in any one of the ξ's, we have a vanishing single transition probability P_{ij} landing in a state in the vanishing part, (7.21). In physical problems, an initial state is after all the final state of another chain of observations. Therefore, we may justifiably exclude states of the vanishing part also as initial states. In any event, symmetry of the physical laws in time (reversibility or inversibility) results in the inverse normalization (7.35), which implies non-existence of vanishing part.

The ergodic theorem is sometimes expressed as a statement regarding the eventual return to the initial state. In actual physical problems in classical physics, a rigorous return to the initial state is not to be expected, but the return to a state infinitely close to the original

state (the so-called quasi-ergodic theorem) is sufficient. However, in our simplified theory, a rigorous return to the initial state can be concluded in classical physics.

Theorem VI. If the values of P_{ij} are limited to zero and unity, then there exists a value of n such that

$$P_{ii}^{(n)} = 1, \tag{7.37}$$

except for S_i belonging to the vanishing part.

Proof: Taking $i = j$ in (7.20), we have

$$\Omega_{ii}^{\infty} > 0 \tag{7.38}$$

Hence, for large enough values of n, we have $\Omega_{ii}^{(n)} > 0$, $(n \geq n_0)$ (7.39)

Comparing (7.39) with (7.9), we see that there must be a value of n (indeed there must be an infinite number of such n's) for which

$$P_{ii}^{(n)} > 0. \tag{7.40}$$

If P_{ij} is zero or unity as we assume, then $P_{ij}^{(n)}$ is also limited to the values zero and unity. Then (7.40) means that there is a value of n for which

$$P_{ii}^{(n)} = 1 \tag{7.41}$$

Taking the smallest value of such n's, we can further infer, in virtue of (7.2),

$$1 = P_{ii}^{(n)} = P_{ii}^{(2n)} = P_{ii}^{(3n)} = \cdots , \tag{7.42}$$

showing a cyclic return to the initial state.* This represents the fundamental fact on which Boltzmann's theorem which Tolman calls "cycle of corresponding collisions" is based.[6] Our proof of the cyclic balance

---

* Taking the smallest common multiple of the n's for various i, (7.37) and (7.42) will become valid for all the i's.

(7.42) is more general than Tolman's argument, since Tolman (i) assumes without verification the ergodic nature of physical phenomena, (ii) utilizes throughout the chaos hypothesis regarding the positions of molecules, and (iii) limits his discussion to collision processes. Our proof admittedly is conditioned by the assumption that the number of possible states is finite.

It is not surprising that the proof of Theorem VI does not utilize the inverse-normalization, since in classical physics, exclusion of vanishing part immediately results in inverse normalization. Indeed, states belonging to $\gamma' = \gamma - \mathfrak{N}$ are connected in this case by a one-to-one correspondence.*

The theorem of cyclic balance (7.42) can be considered as a generalization of the theorem of detailed balance, which is a special case of (7.42) for $n = 2$. Indeed, from

$$P_{ij} = P_{ji}, \tag{7.43}$$

follows

$$P_{ii}^{(2)} = \sum_k P_{ik} P_{ki} = \sum_k (P_{ik})^2. \tag{7.44}$$

In classical physics, only one of $P_{ik} (k = 1,2,\ldots)$ is different from zero and equal to unity. Thence,

$$P_{ii}^{(2)} = 1. \tag{7.45}$$

In quantum physics also, we have (7.40), but it is not of particular interest. Probably another generalization of (7.43) may be more useful.

---

* This means that P_{ij} is actually a permutation, and it is obvious that a finite number of repeated permutations results in the identity transformation. The author's thanks are due to Prof. S. Kakutani for reading this section before publication and for pointing out various interesting facts pertinent to the subject matter, including the point mentioned in this footnote.

In classical as well as quantum physics, we have

$$\Omega_{ij}^{\infty} = \Omega_{ji}^{\infty} \tag{7.46}$$

which is an obvious consequence of Theorem V. This means that the time average of transition probability from S_i to S_j is equal to the time average of transition probability from S_j to S_i. This is also equal to the time average of probability of return to the original state: Ω_{ii}^{∞} or Ω_{jj}^{∞}. Theorem (7.46) may be called theorem of "long-range balance."

In the entire, foregoing discussions, we used chiefly $\Omega_{ij}^{(n)}$ instead of $P_{ij}^{(n)}$, but it is evident that if $P_{ij}^{(n)}$ $(n \to \infty)$ has a limit, this limit is the same as Ω_{ij}^{∞}.

It should also be noticed that we can apply all the foregoing discussions to $W(S_i \to S_j;t)$. Even in classical physics this quantity is not limited to the values zero and unity. Therefore, what have been stated above with regard to quanttum theoretical P's applies, mutatis mutandis, to the W's.

For applications of our results to quantum theoretical problems, the following remarks should be kept in mind, beside the remark we already made with regard to chains of repeated observations in connection with (7.2). If the S's are defined by operators which commute with the exact total Hamiltonian, then $P_{ij} = \delta_{ij}$, and the sub-shell will reduce to one quantum state. In this case, the entire argument loses its physical interest. Therefore, the essential point in the discussion of ergodicity

lies in the tacit assumption that the operators defining the states S_i do not commute with the exact, total Hamiltonian. In fact, this assumption is adopted, explicitly or implicitly, in any version of H-theorem or ergodic theorem in quantum physics.[14] In applications to Thermodynamics, it is necessary to introduce the idea of macroscopic cells on the macroscopically defined energy shell.[14] Our derivation, which does not make use of this concept, should therefore be considered as a simplified model which serves only to clarify the mathematical gist underlying more elaborate formulations.

The author would like to thank Dr. Cecile DeWitt whose instructive seminar talk partially motivated the author to undertake this work.

TA7
.U6
no.2
27444
Watanabe
Symmetry of physical laws, Pt. 1.